

Dicionário - Produtos

João Pedro Magro

Última atualização no dia 05 de junho de 2022 às 22:26

```
#Selecting col's from lilian's dic. and excluding brand names
DicLil <- DicLil|>
  filter(`marca` (1=sim; 0=não)` == "x")|>
  select("y",item, processamento, variedade)

#Changing header names
colnames(CadProdAlim) <- c("CodProdPOF","DescProd")
colnames(DicLil) <- c("RawUsed", "Item", "Proce", "Vari")

#Setting to simple string format
CadProdAlim <- CadProdAlim|>
  mutate(DescProd = tolower(DescProd),
         RawUsedPOF = word(DescProd, 1)) #Spliting indentifyer from IBGE's Dic.

DicLil[DicLil == "x"]<- NA
```

Raw Run

```
#Joining data, by string, max. dist=0
JoinProdAlimTemp <- DicLil|>
  stringdist_left_join(CadProdAlim, by = c("RawUsed"="RawUsedPOF"), max_dist = 0)

JoinProdAlim0 <- JoinProdAlimTemp|>
  filter(!is.na(CodProdPOF))

JoinProdAlimTemp <- JoinProdAlimTemp|>
  filter(is.na(CodProdPOF))

JoinProdAlimTemp <- JoinProdAlimTemp[,1:4]
#Joining data, by string, max. dist=1

JoinProdAlimTemp <- JoinProdAlimTemp|>
  stringdist_left_join(CadProdAlim, by = c("RawUsed"="RawUsedPOF"), max_dist = 1)

JoinProdAlim1 <- JoinProdAlimTemp|>
  filter(!is.na(CodProdPOF))
```

Head dos match

Distância máxima de caracteres = 0

```
kbl(head(JoinProdAlim0),table.attr = "style = \"color: black;\"", position = "c") |>
  kable_styling(font_size = 16, position = "c", full_width = FALSE)
```

RawUsed	Item	Proce	Vari	CodProdPOF	DescProd	RawUsedPOF
abacate	abacate	in natura	NA	6802701	abacate	abacate
abacaxi	abacaxi	in natura	NA	6802601	abacaxi	abacaxi
abobora	abobora	in natura	NA	6703901	abobora	abobora
abobora	abobora	in natura	NA	8578404	abobora com quiabo	abobora

RawUsed	Item	Proce	Vari	CodProdPOF	DescProd	RawUsedPOF
abobrinha	abobrinha	in natura	NA	6703701	abobrinha	abobrinha
acafrao	acafrao	in natura	NA	6401601	acafrao	acafrao

Distância máxima de caracteres = 1

```
kbl(head(JoinProdAlim1),table.attr = "style = \"color: black;\"", position = "c") |>
  kable_styling(font_size = 16, position = "c", full_width = FALSE)
```

RawUsed	Item	Proce	Vari	CodProdPOF	DescProd	RawUsedPOF
alum	NA	NA	NA	7281201	atum fresco	atum
alum	NA	NA	NA	7703402	atum em conserva	atum
alum	NA	NA	NA	7706302	atum em conserva light	atum
americana	alface	in natura	NA	8500308	americano	americano
atc	NA	NA	NA	6803803	ata	ata
azedo	NA	NA	NA	6701904	azeda	azeda